Counting Problems and Semisimple Groups

Alex Eskin[1]

Abstract. Some natural counting problems admit extra symmetries related to actions of Lie groups. For these problems, one can sometimes use ergodic and geometric methods, and in particular the theory of unipotent flows, to obtain asymptotic formulas.

We will present counting problems related to diophantine equations, diophantine inequalities and quantum chaos, and also to the study of billiards on rational polygons.

1991 Mathematics Subject Classification: Primary 11J25, 22E40

1 Counting lattice points on affine homogeneous varieties

In [EMS2], using ergodic properties of subgroup actions on homogeneous spaces of Lie groups, we study asymtotic behaviour of number of lattice points on certain affine varieties. Consider for instance the following:

Let $p(\lambda)$ be a monic polynomial of degree $n \geq 2$ with integer coefficients and irreducible over $\mathbb{Q}$. Let $M_n(\mathbb{Z})$ denote the set of $n \times n$ integer matrices, and put

$$V_p(\mathbb{Z}) = \{A \in M_n(\mathbb{Z}) \ : \ \det(\lambda I - A) = p(\lambda)\}.$$

Hence $V_p(\mathbb{Z})$ is the set of integral matrices with characteristic polynomial $p(\lambda)$. Consider the norm on $n \times n$ real matrices given by $\|(x_{ij})\| = \sqrt{\sum_{ij} x_{ij}^2}$, and let $N(T, V_p)$ denote the number of elements of $V_p(\mathbb{Z})$ with norm less than T.

Theorem 1.1 *Suppose further that $p(\lambda)$ splits over $\mathbb{R}$, and for a root α of $p(\lambda)$ the ring of algebraic integers in $\mathbb{Q}(\alpha)$ is $\mathbb{Z}[\alpha]$. Then, asymptotically as $T \to \infty$,*

$$N(T, V_p) \sim \frac{2^{n-1} h R \omega_n}{\sqrt{D} \cdot \prod_{k=2}^{n} \Lambda(k/2)} T^{n(n-1)/2}$$

where h is the class number of $\mathbb{Z}[\alpha]$, R is the regulator of $\mathbb{Q}(\alpha)$, D is the discriminant of $p(\lambda)$, ω_n is the volume of the unit ball in $\mathbb{R}^{n(n-1)/2}$, and $\Lambda(s) = \pi^{-s}\Gamma(s)\zeta(2s)$.

[1]Supported by the Sloan Foundation and the Packard Foundation

Example 1 is a special case of the following counting problem which was first studied in [DRS] and [EMc].

The counting problem: Let W be a real finite dimensional vector space with a $\mathbb{Q}$ structure and V a Zariski closed real subvariety of W defined over $\mathbb{Q}$. Let G be a reductive real algebraic group defined over $\mathbb{Q}$, which acts on W via a $\mathbb{Q}$-representation $\rho : G \to \mathrm{GL}(W)$. Suppose that G acts transitively on V. Let $\|\cdot\|$ denote a Euclidean norm on W. Let B_T denote the ball of radius $T > 0$ in W around the origin, and define

$$N(T, V) = |V \cap B_T \cap \mathbb{Z}^n|,$$

the number of integral points on V with norm less than T. We are interested in the asymptotics of $N(T, V)$ as $T \to \infty$.

We use the rich theory of unipotent flows on homogeneous spaces developed in [Mar2], [DM1], [Rat1], [Rat2], [Rat3], [Rat4], [Sha1] and [DM3] to obtain results in this direction.

Let Γ be a subgroup of finite index in $G(\mathbb{Z})$ such that $\Gamma W(\mathbb{Z}) \subset W(\mathbb{Z})$. By a theorem of Borel and Harish-Chandra [BH-C], $V(\mathbb{Z})$ is a union of finitely many Γ-orbits. Therefore to compute the asymptotics of $N(T, V)$ it is enough to consider each Γ-orbit, say $\mathcal{O}$, separately and compute the asymptotics of

$$N(T, V, \mathcal{O}) = |\mathcal{O} \cap B_T|.$$

Suppose that $\mathcal{O} = \Gamma v_0$ for some $v_0 \in V(\mathbb{Z})$. Then the stabilizer $H = \{g \in G : gv_0 = v_0\}$ is a reductive real algebraic $\mathbb{Q}$-subgroup, and $V \cong G/H$. Define

$$R_T = \{gH \in G/H : gv_0 \in B_T\},$$

the pullback of the ball B_T to G/H.

Assume that G^0 and H^0 do not admit nontrivial $\mathbb{Q}$-characters. Then by the theorem of Borel and Harish-Chandra, G/Γ admits a G-invariant (Borel) probability measure, say μ_G, and $H/(\Gamma \cap H)$ admits an H-invariant probability measure, say μ_H. Now the natural inclusion $H/(\Gamma \cap H) \hookrightarrow G/\Gamma$ is an H-equivariant proper map. Let $\pi : G \to G/\Gamma$ be the natural quotient map. Then the orbit $\pi(H)$ is closed, $H/(\Gamma \cap H) \cong \pi(H)$, and μ_H can be treated as a measure on G/Γ supported on $\pi(H)$. Such finite invariant measures supported on closed orbits of subgroups are called *homogeneous measures*. Let $\lambda_{G/H}$ denote the (unique) G-invariant measure on G/H induced by the normalization of the Haar measures on G and H.

To state our result in the general setting, we need another definition:

Definition 1.2 Let G and H be as in the counting problem. For a sequence $T_n \to \infty$, the sequence $\{R_{T_n}\}$ of open sets in G/H is said to be *focused*, if there exist a proper connected reductive real algebraic $\mathbb{Q}$-subgroup L of G containing H^0 and a compact set $C \subset G$ such that

$$\limsup_{n\to\infty} \frac{\lambda_{G/H}(q_H(CL(Z(H^0) \cap \Gamma)) \cap R_{T_n})}{\lambda_{G/H}(R_{T_n})} > 0,$$

where $q_H : G \to G/H$ is the natural quotient map.

Our main counting result is the following:

Theorem 1.3 *Let G and H be as in the counting problem. Suppose that H^0 is not contained in any proper $\mathbb{Q}$-parabolic subgroup of G^0 (equivalently, $Z(H)/(Z(H)\cap\Gamma)$ is compact), and for some sequence $T_n \to \infty$ with bounded gaps, the sequence $\{R_{T_n}\}$ is not focused. Then asymptotically*

$$N(T, V, \mathcal{O}) \sim \lambda_{G/H}(R_T).$$

For the case when H is an affine symmetric subgroup of G this result was proved previously in [DRS] using harmonic analysis; subsequently a simpler proof using the mixing property of the geodesic flow appeared in [EMc]. We note that focusing cannot occur for the affine symmetric case.

In general, focusing does not seem to occur for most natural examples, even though it does happen; see [EMS2] for an example.

Translates of homogeneous measures. The following theorem is the main ergodic theoretic result which allows us to investigate the counting problems. The result is also of general interest, especially from the view point of ergodic theory on homogeneous spaces of Lie groups.

Theorem 1.4 *Let G be a connected real algebraic group defined over $\mathbb{Q}$, $\Gamma \subset G(\mathbb{Q})$ an arithmetic lattice in G with respect to the $\mathbb{Q}$-structure on G, and $\pi : G \to G/\Gamma$ the natural quotient map. Let $H \subset G$ be a connected real algebraic $\mathbb{Q}$-subgroup admitting no nontrivial $\mathbb{Q}$-characters. Let μ_H denote the H-invariant probability measure on the closed orbit $\pi(H)$. For a sequence $\{g_i\} \subset G$, suppose that the translated measures $g_i\mu_H$ converge to a probability measure μ on G/Γ. Then there exists a connected real algebraic $\mathbb{Q}$-subgroup L of G containing H such that the following holds:*

(i) There exists $c_0 \in G$ such that μ is a $c_0Lc_0^{-1}$-invariant measure supported on $c_0\pi(L)$.

In particular, μ is a homogeneous measure.

(ii) There exist sequences $\{\gamma_i\} \subset \Gamma$ and $c_i \to c_0$ in G such that $\gamma_i H \gamma_i^{-1} \subset L$ and $g_iH = c_i\gamma_iH$ for all but finitely many i.

In order to be able to apply Theorem 1.4 to the problem of counting, we need to know some conditions under which the sequence $\{g_i\mu_H\}$ of probability measures does not escape to infinity. A neccesary and sufficient condition on the reductive $\mathbb{Q}$-group H is given in [EMS1].

The proof of Theorem 1.4 is based on Ratner's measure classification theorem. A key observation is that the limit measure μ is invariant under some unipotent element. Still the result does not follow immediately from Ratner's theorem since we do not know that μ is ergodic. In the case when H is itself generated by unipotent elements, the proof is simpler: see [MS].

Correspondence between counting and translates of measures. We recall some observations from [DRS, Sect. 2]; see also [EMc]. Let the notation be as in the counting problem stated in the introduction. For $T > 0$, define a function F_T on G by

$$F_T(g) = \sum_{\gamma \in \Gamma/(H\cap\Gamma)} \chi_T(g\gamma v_0),$$

where χ_T is the characteristic function of B_T. By construction F_T is left Γ-invariant, and hence it will be treated as a function on G/Γ. Note that

$$F_T(e) = \sum_{\gamma \in \Gamma/(H\cap\Gamma)} \chi_T(\gamma v_0) = N(T, V, \mathcal{O}).$$

Since we expect, as in Theorem 1.3, that $N(T, V, \mathcal{O}) \sim \lambda_{G/H}(R_T)$, we define $\hat{F}_T(g) = \frac{1}{\lambda_{G/H}(R_T)} F_T(g)$. Thus Theorem 1.3 is the assertion $\hat{F}_T(e) \to 1$ as $T \to \infty$.

The connection between Theorem 1.3 and Theorem 1.4 is via the following formula (see [DRS] and [EMc]):

$$\langle \hat{F}_T, \psi \rangle = \frac{1}{\lambda_{G/H}(R_T)} \int_{R_T} \left(\int_{G/\Gamma} \overline{\psi}\, d(g\mu_H) \right) d\lambda_{G/H}(g),$$

where ψ is any function in $C_0(G/\Gamma)$ and $g\mu_H$ is the translated measure as in Theorem 1.4.

If the non-focusing assumption is satisfied, then by Theorem 1.4, for "most" values of g, the inner integral will approach $\int_{G/\Gamma} \overline{\psi}\, d\mu = \langle 1, \psi \rangle$. Thus, $\hat{F}_T \to 1$ in the weak-star topology on $L^\infty(G/\Gamma, \mu_G)$. In can then be shown that $\hat{F}_T \to 1$ uniformly on compact sets.

2 A quantitative version of the Oppenheim conjecture

Let Q be an indefinite nondegenerate quadratic form in n variables. Let $\mathcal{L}_Q = Q(\mathbb{Z}^n)$ denote the set of values of Q at integral points. The Oppenheim conjecture, proved by Margulis (cf. [Mar2]) states that if $n \geq 3$, and Q is not proportional to a form with rational coefficients, then $\mathcal{L}_Q$ is dense. In joint work with G. Margulis and S. Mozes ([EMM1]) we study some finer questions related to the distribution the values of Q at integral points.

Let ν be a continuous positive function on the sphere $\{v \in \mathbb{R}^n \mid \|v\| = 1\}$, and let $\Omega = \{v \in \mathbb{R}^n \mid \|v\| < \nu(v/\|v\|)\}$. We denote by $T\Omega$ the dilate of Ω by T. Define the following set:

$$V^Q_{(a,b)}(\mathbb{R}) = \{x \in \mathbb{R}^n \mid a < Q(x) < b\}$$

We shall use $V_{(a,b)} = V^Q_{(a,b)}$ when there is no confusion about the form Q. Also let $V_{(a,b)}(\mathbb{Z}) = V^Q_{(a,b)}(\mathbb{Z}) = \{x \in \mathbb{Z}^n \mid a < Q(x) < b\}$. The set $T\Omega \cap \mathbb{Z}^n$ consists of $O(T^n)$ points, $Q(T\Omega \cap \mathbb{Z}^n)$ is contained in an interval of the form $[-\mu T^2, \mu T^2]$, where $\mu > 0$ is a constant depending on Q and Ω. Thus one might expect that for any interval $[a, b]$, as $T \to \infty$,

$$|V_{(a,b)}(\mathbb{Z}) \cap T\Omega| \sim c_{Q,\Omega}(b-a)T^{n-2} \tag{1}$$

where $c_{Q,\Omega}$ is a constant depending on Q and Ω. This may be interpreted as "uniform distribution" of the sets $Q(\mathbb{Z}^n \cap T\Omega)$ in the real line. Our main result is that (1) holds if Q is not proportional to a rational form, and has signature (p, q) with $p \geq 3$, $q \geq 1$. We also determine the constant $c_{Q,\Omega}$.

If Q is an indefinite quadratic form in n variables, Ω is as above and (a, b) is an interval, we show that there exists a constant $\lambda = \lambda_{Q,\Omega}$ so that as $T \to \infty$,

$$\mathrm{Vol}(V_{(a,b)}(\mathbb{R}) \cap T\Omega) \sim \lambda_{Q,\Omega}(b-a)T^{n-2} \tag{2}$$

Our main result is the following:

Theorem 2.1 *Let Q be an indefinite quadratic form of signature (p, q), with $p \geq 3$ and $q \geq 1$. Suppose Q is not proportional to a rational form. Then for any interval (a, b), as $T \to \infty$,*

$$|V_{(a,b)}(\mathbb{Z}) \cap T\Omega| \sim \lambda_{Q,\Omega}(b-a)T^{n-2}$$

where $n = p + q$, and $\lambda_{Q,\Omega}$ is as in (2).

Only the upper bound in this formula is new: the asymptotically exact lower bound was proved in [DM3]. Also a lower bound with a smaller constant was obtained independently by M. Ratner, and by S. G. Dani jointly with S. Mozes (both unpublished).

If the signature of Q is $(2, 1)$ or $(2, 2)$ then no universal formula like (1) holds. In fact, we have the following theorem:

Theorem 2.2 *Let Ω_0 be the unit ball, and let $q = 1$ or 2. Then for every $\epsilon > 0$ and every interval (a, b) there exists a quadratic form Q of signature $(2, q)$ not proportional to a rational form, and a constant $c > 0$ such that for an infinite sequence $T_j \to \infty$,*

$$|V_{(a,b)}(\mathbb{Z}) \cap T\Omega_0| > cT_j^q(\log T_j)^{1-\epsilon}.$$

The case $q = 1$, $b \leq 0$ of Theorem 2.2 was noticed by P. Sarnak and worked out in detail in [Bre]. The quadratic forms constructed are of the form $x_1^2 + x_2^2 - \alpha x_3^2$, or $x_1^2 + x_2^2 - \alpha(x_3^2 + x_4^2)$, where α is extremely well approximated by squares of rational numbers.

However in the $(2, 1)$ and $(2, 2)$ cases, we can still establish an upper bound of the form $cT^q \log T$. This upper bound is effective, and is uniform over compact sets in the set of quadratic forms. We also give an effective uniform upper bound for the case $p \geq 3$.

Theorem 2.3 *Let $\mathcal{O}(p, q)$ denote the space of quadratic forms of signature (p, q) and discriminant ± 1, let $n = p + q$, (a, b) be an interval, and let $\mathcal{D}$ be a compact* *subset of $\mathcal{O}(p,q)$. Let ν be a continuous positive function on the unit sphere and let $\Omega = \{v \in \mathbb{R}^n \mid \|v\| < \nu(v/\|v\|)\}$. Then, if $p \geq 3$ there exists a constant c depending only on $\mathcal{D}$, (a,b) and Ω such that for any $Q \in \mathcal{D}$ and all $T > 1$,*

$$|V_{(a,b)}(\mathbb{Z}) \cap T\Omega| < cT^{n-2}$$

If $p = 2$ and $q = 1$ or $q = 2$, then there exists a constant $c > 0$ depending only on $\mathcal{D}$, (a,b) and Ω such that for any $Q \in \mathcal{D}$ and all $T > 2$,

$$|V_{(a,b)} \cap T\Omega \cap \mathbb{Z}^n| < cT^{n-2}\log T$$

Also, for the $(2,1)$ and $(2,2)$ cases, we have the following "almost everywhere" result:

THEOREM 2.4 *For almost all quadratic forms Q of signature $(p,q) = (2,1)$ or $(2,2)$*

$$|V_{(a,b)}(\mathbb{Z}) \cap T\Omega| \sim \lambda_{Q,\Omega}(b-a)T^{n-2}$$

where $n = p + q$, and $\lambda_{Q,\Omega}$ is as in (2).

CONNECTION WITH QUANTUM CHAOS. It has been suggested by Berry and Tabor that the distribution of the local spacings between eigenvalues of the quantization of a completely integrable Hamiltonian is Poisson. For the Hamiltonian which is the geodesic flow on the flat 2-torus, it was noted by P. Sarnak [Sar] that this problem translates to one of the spacing between the values at integers of a binary quadratic form, and is related to the quantitative Oppenheim problem. We briefly recall the connection following [Sar].

Let β^2 be a positive irrational number, and let M_β denote the rectangular torus with the flat metric and sides π and π/β. Let Λ_β denote the spectrum of the Laplace operator on M_β, i.e.

$$\Lambda_\beta = \{P_\beta(m,n) \ : \ m,n \geq 0, \text{ or } m = 0, n \geq 0, \quad m,n \in \mathbb{Z}\}$$

where $P_\beta(x,y)$ denotes the positive definite quadratic form $x^2 + \beta^2 y^2$. We label the elements of Λ_β (with multiplicity) by

$$0 = \lambda_0(\beta) < \lambda_1(\beta) \leq \lambda_2(\beta) \ldots$$

It is easy to see that Weyl's law holds, i.e.

$$|\{j \ : \ \lambda_j(\beta) \leq T\}| \sim c_\beta T$$

where $c_\beta = \pi/(4\beta)$ is related to the area of M_β. We are interested in the distribution of the local spacings $\lambda_j(\beta) - \lambda_k(\beta)$. In particular, set

$$R_\beta(a,b,T) = \frac{|\{(j,k) \ : \ \lambda_j(\beta) \leq T, \lambda_k(\beta) \leq T, j \neq k, a \leq \lambda_j(\beta) - \lambda_k(\beta) \leq b\}|}{T}$$

The quantity R_β is called the pair correlation. The random number (Poisson) model predicts that

$$\lim_{T\to\infty} R_\beta(a,b,T) = c_\beta^2(b-a). \tag{3}$$

Note that the differences $\lambda_j(\beta) - \lambda_k(\beta)$ are precisely the integral values of the quadratic form $Q_\beta(x_1,x_2,x_3,x_4) = x_1^2 - x_3^2 + \beta^2(x_2^2 - x_4^2)$.

P. Sarnak considered in [Sar] a two-parameter family of flat 2-tori and showed that (3) holds on a set of full measure of these tori. Some similar results for forms of higher degree were proved in [Va1] and [Va2].

These methods, however, cannot be used to construct a specific torus for which (3) holds. In [EMM2], using a refinement of the methods of [EMM1] we establish (3) under a mild diophantine condition on β, and in particular for any irrational algebraic β. Our main result is the following:

Theorem 2.5 *Suppose β^2 is* diophantine, *i.e. there exists $N > 0$ such that for all relatively prime pairs of integers (p,q), $|\beta^2 - p/q| > q^{-N}$. Then, for any interval (a,b), (3) holds, i.e.*

$$\lim_{T\to\infty} R_\beta(a,b,T) = c_\beta^2(b-a).$$

In particular, the set of $\beta \subset \mathbb{R}$ for which (3) does not hold has zero Hausdorff dimension.

Thus, if β^2 is diophantine, then M_β has a spectrum whose pair correlation satisfies the Berry-Tabor conjecture.

We note that some diophantine condition in Theorem 2.5 is needed in view of Theorem 2.2.

Quadratic Forms. We now relate the counting problem of Theorem 2.1 to a certain integral expression involving the orthogonal group of the quadratic form and the space of lattices G/Γ, where $G = SL(n,\mathbb{R})$, $\Gamma = SL(n,\mathbb{Z})$. Let f be a bounded function on $\mathbb{R}^n - \{0\}$ vanishing outside a compact subset. For a lattice $\Delta \in SL(n,\mathbb{R})$ let

$$\tilde{f}(\Delta) = \sum_{v\in\Delta} f(v) \tag{4}$$

Let $n \geq 3$, and let $p \geq 2$. We denote $n-p$ by q, and assume $q > 0$. Let $\{e_1, e_2, \dots e_n\}$ be the standard basis of $\mathbb{R}^n$. Let Q_0 be the quadratic form defined by

$$Q_0\left(\sum_{i=1}^{n} v_i e_i\right) = 2v_1 v_n + \sum_{i=2}^{p} v_i^2 - \sum_{i=p+1}^{n-1} v_i^2 \quad \text{for all } v_1,\dots,v_n \in \mathbb{R}.$$

It is straightforward to verify that Q_0 has signature (p,q). For each quadratic form Q and $g \in G$, let Q^g denote the quadratic form defined by $Q^g(v) = Q(gv)$ for all $v \in \mathbb{R}^n$. By the well known classification of quadratic forms over $\mathbb{R}$, for each $Q \in \mathcal{O}(p,q)$ there exists $g \in G$ such that $Q = Q_0^g$. For any quadratic form Q let $SO(Q)$ denote the special orthogonal group corresponding to Q; namely $\{g \in G \mid Q^g = Q\}$. Let $H = SO(Q_0)$. Then the map $H \backslash G \to \mathcal{O}(p,q)$ given by $Hg \to Q_0^g$ is a homeomorphism. If $Q = Q_0^g$, let Δ_Q denote the lattice $g\mathbb{Z}^n$.

For $t \in \mathbb{R}$, let a_t be the linear map so that $a_t e_1 = e^{-t} e_1$, $a_t e_n = e^t e_n$, and $a_t e_i = e_i$, $2 \le i \le n-1$. Then the one-parameter group $\{a_t\}$ is contained in H. Let $\hat{K}$ be the subgroup of G consisting of orthogonal matrices, and let $K = H \cap \hat{K}$. It is easy to check that K is a maximal compact subgroup of H, and consists of all $h \in H$ leaving invariant the subspace spanned by $\{e_1 + e_n, e_2, \ldots, e_p\}$. We denote by m the normalized Haar measure on K.

Suppose for simplicity that the set Ω in Theorem 2.1 is invariant under the action of K. Then, it can be shown that for a suitably chosen function f on $\mathbb{R}^n$, $|V_{(a,b)}(\mathbb{Z}) \cap T\Omega|$ can be well approximated by the following expression:

$$T^{n-2} \int_K \tilde{f}(a_t k \Delta_Q)\, dm(k) \tag{5}$$

where $t = \log T$, and $\tilde{f}$ is as in (4). Thus, Theorem 2.1 can be deduced from the following theorem:

Theorem 2.6 *Suppose $p \ge 3$, $q \ge 1$. Let f be a continuous function on $\mathbb{R}^n$ vanishing outside a compact set. Let $\Delta \in G/\Gamma$ be a unimodular lattice such that $H\Delta$ is not closed. Then*

$$\lim_{t \to +\infty} \int_K \tilde{f}(a_t k \Delta)\, dm(k) = \int_{G/\Gamma} \tilde{f}(y)\, d\mu(y). \tag{6}$$

If in Theorem 2.6, in place of the function $\tilde{f}$ we considered any bounded continuous function ϕ, then (6) would follow easily from [DM3, Theorem 3]). This theorem is a refined version of Ratner's uniform distribution theorem [Rat4]; the proof uses Ratner's measure classification theorem (see [Rat1, Rat2, Rat3]), Dani's theorem on the behavior of unipotent orbits at infinity [Dan1, Dan2], and "linearization" techniques.

Both [DM3, Theorem 3] and Ratner's uniform distribution theorem hold for bounded continuous functions, but not for arbitrary continuous functions from $L^1(G/\Gamma)$. However, for a non-negative bounded continuous function f on $\mathbb{R}^n$, the function $\tilde{f}$ defined in (4) is non-negative, continuous, and L^1 but unbounded (it is in $L^s(G/\Gamma)$ for $1 \le s < n$, where $G = SL(n,\mathbb{R})$, and $\Gamma = SL(n,\mathbb{Z})$). As it was done in [DM3] it is possible to obtain asymptotically exact lower bounds by considering bounded continuous functions $\phi \le \tilde{f}$. However, to carry out the integral in (5) and prove the upper bounds in the theorems stated above we need to examine carefully the situation at the "cusp" of G/Γ, i.e outside of compact sets. Some techniques for handling this were developed in [Mar1], [Dan1], [Dan2]; see also [KM] for a simplified proof and some interesting applications to the metric theory of diophantine approximations. However, these techniques are not sufficient for this problem.

Let Δ be a lattice in $\mathbb{R}^n$. We say that a subspace L of $\mathbb{R}^n$ is Δ-*rational* if $L \cap \Delta$ is a lattice in L. For any Δ-rational subspace L, we denote by $d_\Delta(L)$ or simply by $d(L)$ the volume of $L/(L \cap \Delta)$. Let us note that $d(L)$ is equal to the norm of $e_1 \wedge \cdots \wedge e_\ell$ in the exterior power $\bigwedge^\ell(\mathbb{R}^n)$ where $\ell = \dim L$ and $(e_1, \cdots, e_\ell)$ is a basis over $\mathbb{Z}$ of $L \cap \Delta$. If $L = \{0\}$ we write $d(L) = 1$. A lattice is Δ *unimodular* if $d_\Delta(\mathbb{R}^n) = 1$. The space of unimodular lattices is isomorphic to $SL(n, \mathbb{R})/SL(n, \mathbb{Z})$.

Let us introduce the following notation:

$$\alpha_i(\Delta) = \sup\left\{ \frac{1}{d(L)} \,\middle|\, L \text{ is a } \Delta\text{-rational subspace of dimension } i \right\}, \quad 0 \le i \le n,$$
$$\alpha(\Delta) = \max_{0 \le i \le n} \alpha_i(\Delta).$$

The following lemma is known as the "Lipshitz Principle":

Lemma 2.7 ([Sch, Lemma 2]) *Let f be a bounded function on $\mathbb{R}^n$ vanishing outside a compact subset. Then there exists a positive constant $c = c(f)$ such that*

$$\tilde{f}(\Delta) < c\alpha(\Delta) \tag{7}$$

for any lattice Δ in $\mathbb{R}^n$. Here $\tilde{f}$ is the function on the space of lattices defined in (4).

By (7) the function $\tilde{f}(g)$ on the space of unimodular lattices G/Γ is majorized by the function $\alpha(g)$. The function α is more convenient since it is invariant under the left action of the maximal compact subgroup $\hat{K}$ of G, and its growth rate at infinity is known explicitly. Theorem 2.6 is proved by combining [DM3, Theorem 3] with the following integrability estimate:

Theorem 2.8 *If $p \ge 3$, $q \ge 1$ and $0 < s < 2$, or if $p = 2$, $q \ge 1$ and $0 < s < 1$, then for any lattice Δ in $\mathbb{R}^n$*

$$\sup_{t>0} \int_K \alpha(a_t k \Delta)^s \, dm(k) < \infty.$$

The upper bound is uniform as Δ varies over compact sets in the space of lattices.

This result can be interpreted as follows. For a lattice Δ in G/Γ and for $h \in H$, let $f(h) = \alpha(h\Delta)$. Since α is left-$\hat{K}$ invariant, f is a function on the symmetric space $X = K \backslash H$. Theorem 2.8 is the statement that if if $p \ge 3$, then the averages of f^s, $0 < s < 2$ over the sets $K a_t K$ in X remain bounded as $t \to \infty$, and the bound is uniform as one varies the base point Δ over compact sets. We remark that in the case $q = 1$, the rank of X is 1, and the sets $K a_t K$ are metric spheres of radius t, centered at the origin.

If $(p, q) = (2, 1)$ or $(2, 2)$, Theorem 2.8 does not hold even for $s = 1$. The following result is, in general, best possible:

Theorem 2.9 *If $p = 2$ and $q = 2$, or if $p = 2$ and $q = 1$, then for any lattice Δ in $\mathbb{R}^n$,*

$$\sup_{t>1} \frac{1}{t} \int_K \alpha(a_t k \Delta) \, dm(k) < \infty,$$

The upper bound is uniform as Δ varies over compact sets in the space of lattices.

We now outline the proof of Theorems 2.8 and 2.9. From its definition, the function $\alpha(g)$ is the maximum over $1 \le i \le n$ of left-$\hat{K}$ invariant functions $\alpha_i(g)$. The main idea of the proof is to show that the α_i satisfy a system of integral inequalities which imply the desired bound.

If $p \ge 3$ and $0 < s < 2$, or if $(p,q) = (2,1)$ or $(2,2)$ and $0 < s < 1$, we show that for any $c > 0$ there exist $t > 0$, and $\omega > 1$ so that the the functions α_i^s satisfy the following system of integral inequalities in the space of lattices:

$$A_t \alpha_i^s \le c_i \alpha_i^s + \omega^2 \max_{0 \le j \le n-i,i} \sqrt{\alpha_{i+j}^s \alpha_{i-j}^s} \tag{8}$$

where A_t is the averaging operator $(A_t f)(\Delta) = \int_K f(a_t k \Delta)$, and $c_i \le c$. If $(p,q) = (2,1)$ or $(2,2)$ and $s = 1$, then (8) also holds (for suitably modified functions α_i), but some of the constants c_i cannot be made smaller than 1.

Let $f_i(h) = \alpha_i(h\Delta)$, so that each f_i is a function on the symmetric space X. When one restricts to an orbit of H, (8) becomes:

$$A_t f_i^s \le c_i f_i^s + \omega^2 \max_{0 \le j \le n-i,i} \sqrt{f_{i+j}^s f_{i-j}^s} \tag{9}$$

If $\operatorname{rank} X = 1$, then $(A_t f)(h)$ can be interpreted as the average of f over the sphere of radius t in X, centered at h. We show that if the f_i satisfy (9) then for any $\epsilon > 0$, the function $f = f_{\epsilon,s} = \sum_{0 \le i \le n} \epsilon^{i(n-i)} f_i^s$ satisfies the scalar inequality:

$$A_t f \le cf + b \tag{10}$$

where t, c and b are constants. We show that if c is sufficiently small, then (10) for a fixed t together with the uniform continuity of $\log f$ imply that $(A_r f)(1)$ is bounded as a function of r, which is the conclusion of Theorem 2.8. If $c = 1$, which will occur in the $SO(2,1)$ and $SO(2,2)$ cases, then (10) implies that $(A_r f)(1)$ is growing at most linearly with the radius, which is the conclusion of Theorem 2.9.

3 Actions on the space of quadratic differentials

Given a Riemann surface structure on a closed surface of genus $g > 1$, recall that a holomorphic quadratic differential ϕ is a tensor of the form $\phi(z)dz^2$ in local coordinates with ϕ holomorphic. Away from the zeroes, a coordinate ζ can be chosen so that $\phi = d\zeta^2$, which determines a Euclidean metric $|d\zeta^2|$ in that chart. The change of coordinates away from the zeroes of ϕ are of the form $\zeta \to \pm\zeta + c$, which preserves the Euclidan metric. Consequently, quadratic differentials are sometimes refered to as translation surfaces or flat structures. If one can always take the $+$ sign in the change of coordinates, the translation surface is orientable. Equivalently, the quadratic differential is the square of an abelian differential. We will henceforth adopt the notation S to refer to the structure of a quadratic differential. A zero of order $k \ge 1$ of S defines a cone angle singularity of the metric: there is a neighborhood of the zero such that the metric is of the form

$$ds^2 = dr^2 + ((k+2)r d\theta/2)^2$$

and the cone angle is $(k+2)\pi$. The number of zeroes counting multiplicity is $4g-4$. Let P be a partition of $4g-4$ (i.e. a representaion of $4g-4$ as a sum of positive integers).

Let $QD(g,P)$ denote the set of flat structures S on a surface of genus g whose zero set is given by P; the space $QD(g,P)$ is called a stratum. The term is justified by the fact that the space of all quadratic differentials on Riemann surfaces of genus g is stratified by the spaces $QD(g,P)$ as P varies over the partitions of $4g-4$. The spaces $QD(g,P)$ have projection maps to the Teichmuller space T_g of genus g.

There is an $SL(2,\mathbb{R})$ action on $QD(g,P)$. In each coordinate patch $\zeta \in \mathbb{R}^2$, and for $A \in SL(2,\mathbb{R})$, the action is the linear action $\zeta \to A\zeta$. The fact that the change coordinates is given by $\zeta \to \pm\zeta + c$ says that this is well-defined. This action preserves a measure μ_0 on $QD(g,P)$ ([Mas1], [Ve1]).

A saddle connection of S is a geodesic segment joining two zeroes of S which has no zeroes in its interior. A saddle connection determines a vector in $\mathbb{R}^2$ since in each coordinate chart it is geodesic with respect to Euclidean metric. A closed geodesic that does not pass through any zeroes determines a cylinder of parallel freely homotopic closed geodesics of the same length. Each boundary component of the cylinder consists of one or more parallel saddle connections.

We note that there is a well known construction which associates a surface with a flat structure to each rational polygon (see [ZK], [Gut], [KMS]). In particular counting families of periodic trajectories of the billiard in the polygon is equivalent to counting cylinders of closed geodesics on the surface constructed from the polygon.

In a recent paper [Ve2] Veech observed that this counting problem is analogous to the counting problem of section §2. The results of this section, which are joint work with H. Masur, are inspired by this paper. We now recall the construction of [Ve2]. In the Teichmuller space situation, for a bounded function f on $\mathbb{R}^2 - \{0\}$, Veech defines a function on the moduli space of quadratic differentials by

$$\tilde{f}(S) = \sum_{v \in V(S))} f(v) \tag{11}$$

where for a point x in the moduli space(i.e a surface and a quadratic differential), $V(S)$ denotes the set of vectors in $\mathbb{R}^2$ corresponding to cylinders of closed geodesics on S.

Let $N(S,T)$ denote the number of cylinders periodic geodesics on S of length at most T. In [Ve2] it is shown that for an appropriate function f, $N(S,T)$ is well approximated by an expression of the form:

$$T^2 \int_K \tilde{f}(a_t k \cdot S)\, dm(k) \tag{12}$$

where a_t and k are as in §2 and $t = \log T$. Thus, this problem is remarkably similar to the problem in §2. However since we are not in the homogeneous space setting, we expect that obtaining results as sharp as in §2 would be very difficult. However, some of the techniques mentioned in §2, in particular, the idea of obtaining upper bounds via systems of inequalities can be transferred to this situation. This yields a simplified proof of the quadratic upper bound of [Mas2].

One can also obtain an individual ergodic theorem analogous to Theorem 2.4, using an ergodic theorem due to A. Nevo. More precisely, one can prove the following (in weaker form this theorem is proved in [Ve2]):

Theorem 3.1 *For a (nonorientable) translation surface S and $T > 0$, let $N(S, T)$ denote the number of cylinders of closed geodesics on S such that the norm of the associated vector is at most T. For any P there exists a constant c_P such that for almost all $S \in QD(g, P)$,*

$$N(S, T) \sim c_P T^2$$

In some special examples, one can use the full theory of unipotent flows in a way exactly analogous to that of §2. In particular, one can prove the following:

Theorem 3.2 *For any rational $0 < p/q < 1$, and any α, $0 < \alpha < 1$, let $P = P_{p/q,\alpha}$ denote the polygon whose boundary is the boundary of the unit square $\partial([0,1] \times [0,1])$ union the segment $\{p/q\} \times [0, \alpha]$. Then, for any irrational α,*

$$N(P, T) \sim c_{p/q} T^2$$

where $c_{p/q}$ is some explicitly computable constant.

We note that the case $p/q = 1/2$ is elementary, and was done previously by E. Gutkin and C. Judge (private communication).

References

[BH-C] A. Borel and Harish-Chandra. Arithmetic subgroups of algebraic groups. *Annals of Math* 75 (1962), 485–535.

[BR] M. Borovoi and Z. Rudnick. Hardy–Littlewood varieties and semisimple groups. *Inv. Math.* (1994)

[Bre] T. Brennan. . Princeton University undergraduate thesis, 1994.

[Dan1] S.G. Dani. On orbits of unipotent flows on homogeneous spaces. *Ergod. Theor. Dynam. Syst.* 4(1984), 25–34.

[Dan2] S.G. Dani. On orbits of unipotent flows on homogenous spaces II. *Ergod. Theor. Dynam. Syst* 6(1986), 167–182.

[Dan3] S.G. Dani. Flows on homogeneous spaces and diophantine approximation. In: *Proc. of ICM (1994)*, Zurich.

[DM1] S.G. Dani and G.A. Margulis. Orbit closures of generic unipotent flows on homogeneous spaces of $\mathrm{SL}(3, \mathbb{R})$. *Math.Ann.* 286 (1990), 101–128.

[DM2] S.G. Dani and G.A. Margulis. Asymptotic behaviour of trajectories of unipotent flows on homogeneous spaces. *Indian. Acad. Sci. J.* 101 (1991), 1–17.

[DM3] S.G. Dani and G.A. Margulis. Limit distributions of orbits of unipotent flows and values of quadratic forms. *Advances in Soviet Math.* 16 (1993), 91–137.

[DRS] W. Duke, Z. Rudnick, and P. Sarnak. Density of integer points on affine homogeneous varieties. *Duke Math. J.* 71(1993), 143–180.

[EMc] A. Eskin and C. McMullen. Mixing, counting and equidistribution in Lie groups. *Duke Math. J.* 71(1993), 143–180.

[EMM1] A. Eskin, G. Margulis, and S. Mozes. Upper bounds and asymptotics in a quantitative version of the Oppenheim conjecture. *Ann. of Math. (2)* 147(1998), no. 1, 93–141

[EMM2] A. Eskin, G. Margulis, and S. Mozes. Quadratic forms of signature $(2, 2)$ and eigenvalue spacings on rectangular 2-tori. *Preprint.*

[EMS1] A. Eskin, S. Mozes, and N. Shah. Non-divergence of translates of certain algebraic measures. *Geom. Funct. Anal.* 7(1997), no. 1, 48–80.

[EMS2] Alex Eskin, Shahar Mozes, and Nimish Shah. Unipotent flows and counting lattice points on homogeneous varieties. *Annals of Math.* (1996), 253–299.

[Gut] E. Gutkin. Billiards on almost integrable polyhedral surfaces. *Ergodic Th. Dyn Syst.* 4 (1984) pp. 569–584

[KM] D. Kleinbok and G. Margulis. Flows on homogeneous spaces and Diophantine approximation on manofolds. *To appear in Annals of Math.*

[KMS] S. Kerckhoff, H. Masur and J. Smillie, *Ergodicity of Billiard flows and quadratic differentials*, Annals of Math. 124 (1986), 293–311.

[Marklof] J. Marklof. Spectral form factors of rectangular billiards. *Preprint.*

[Mar1] G. A. Margulis. On the action of unipotent groups in the space of lattices. In *Lie Groups and their representations, (Proc. of Summer School in Group Representations, Bolyai Janos Math. Soc., Akademai Kiado, Budapest, 1971)*, pages 365–370. Halsted, New York, 1975.

[Mar2] G.A. Margulis. Discrete subgroups and ergodic theory. In: *Number Theory, Trace Formulas and Discrete Groups* (symposium in honour of A. Selberg, Oslo). Academic Press, 1989, pp. 377-398.

[Mar3] G.A. Margulis. Dynamical and ergodic properties of subgroups actions on homogeneous spaces with applications to number theory. In:*Proc. of ICM* (Kyoto, 1990). Math. Soc. of Japan and Springer, 1991, pp. 193-215.

[Mar4] G.A. Margulis. Oppenheim Conjecture. In: *Fields Medalists Lectures,* World Scientific (1997), pp. 272–327.

[MS] S. Mozes and N.A. Shah. On the space of ergodic invariant measures of unipotent flows. *Ergodic Th. and Dynam. Syst.* 15 (1995), 149–159.

[Mas1] H. Masur. Interval exchange transformations and measured foliations. *Annals of Math.* 115(1982), 169–200.

[Mas2] H. Masur. The growth rate of trajectories of a quadratic differential. *Ergodic Th. and Dynam. Syst.* 10 (1990), 151–176.

[Rat1] M. Ratner. Strict measure rigidity for nilpotent subgroups of solvable groups. *Invent. Math.* 101 (1990), 449–482.

[Rat2] M. Ratner. On measure rigidity of unipotent subgroups of semisimple groups. *Acta. Math.* 165 (1990), 229–309.

[Rat3] M. Ratner. On Ragunathan's measure conjecture. *Annals of Math.* 134 (1991), 545–607.

[Rat4] M. Ratner. Raghunathan's topological conjecture and distributions of unipotent flows. *Duke Math. J.* 63 (1991), 235–290.

[Rat5] M. Ratner. Interactions between Lie groups, ergodic theory and number theory. *Proc. of ICM (Zurich, 1994).*

[Sar] P. Sarnak. Values at integers of binary quadratic forms. Preprint.

[Sch] W. Schmidt. Asymptotic formulae for point lattices of bounded determinant and subspaces of bounded height. *Duke Math. J.* 35(1968), 327–339.

[Sha1] N. A. Shah. Uniformly distributed orbits of certain flows on homogeneous spaces. *Math. Ann.* 289 (1991), 315–334.

[Ve1] W. Veech. Gauss measures for transformations on the space of interval exchange maps. *Ann. of Math.* (2) 115 (1982), no. 1, 201–242.

[Ve2] W. Veech, *Siegel measures*, preprint.

[Va1] J. Vanderkam. Values at integers of homogeneous polynomials. *To appear in the Duke Math. J.*

[Va2] J. Vanderkam. Pair correlation of four-dimensional flat tori. *To appear in the Duke Math. J.*

[ZK] A.N. Zemlyakov, A.B. Katok. Topological transitivity of billiards in polygons. *Matem. Zametki* 18(2) (1975) pp. 291–300. English translation in *Math. Notes* 18(2) (1976) pp. 760–764.

Alex Eskin
Department of Mathematics
University of Chicago
Chicago
Illinois, USA
eskin@math.uchicago.edu